Technical Disclosure Commons

Defensive Publications Series

July 03, 2017

Quality Assessment Signals

John D. Lanza
Foley & Lardner LLP

John D. Lanza
Foley & Lardner LLP

QUALITY ASSESSMENT SIGNALS

ABSTRACT

Systems and methods described herein allow for monitoring and adjustment of the quality of conversations between users and a virtual personal assistant. A data processing system hosting the virtual personal assistant can include a feedback monitoring component configured to determine parameters related to (or characteristics of) communication links between the data processing system and the client device, speech recognition and natural language processing of received audio signals, and/or user online activities responsive to content from the data processing system presented to the user. The data processing system can determine one or more quality metrics based on the collected parameters or characteristics, and adjust one or more content selection process parameters based on the determined quality signals.

DETAILED DESCRIPTION

In a voice-based interaction computer environment, users can engage in audio conversations with respective client devices. For instance, a client device implementing a front-end module of a virtual personal assistant can allow for two-way conversations with the respective user. Such conversations usually relate to user requests to be fulfilled by a data processing system remotely connected to a client device of the user. The voice-based interaction, even though convenient and attractive to users, can pose technical challenges with respect to achieving and maintaining a good or at least satisfactory conversation quality to guarantee a good user experience. In particular, the quality of the audio signals exchanged between the client device and the data processing system, the quality of interpreting audio

signals received from the client device, and/or the quality of content presented to the user can affected by various factors, such as ambient noise, characteristics of communication links, characteristics of user's voice, parameters of software modules or algorithms used to handle conversations with users, or a combination thereof.

The present disclosure is generally directed to using audio data received from client devices associated with a user to assess the quality of the conversation with user and adjust one or more parameters of a content selection process based on the determined quality. The data processing system can collect a variety of parameters related to (or characteristics of) communication links between the data processing system and the client device, speech recognition and natural language processing of received audio signals, and/or user online activities responsive to content from the data processing system presented to the user. The data processing system can determine one or more quality signals (or quality metrics) based on the collected parameters or characteristics, and adjust one or more content selection process parameters based on the determined quality signals.

FIG. 1 is flowchart illustrating an example method 100 for monitoring and adjusting the quality of a conversation in a voice-based interaction computer environment. The method 100 can be performed by the data processing system hosting the backend of a virtual personal assistant and serving client devices with requested content. At step 105 the method 100 can include identifying a user request based on a received input audio signal. The input audio signal can be received by a client device and can be indicative of a request or command made by a corresponding user. The user can initiate a conversation with an instance of the virtual personal assistant (e.g., Google Assistant). The user can make a request for an online service or action (e.g., an audio search query, making or checking an airline flight reservation, online purchase of

movie tickets, rendering of an online live stream, scheduling of a cab service, etc.). The client device can receive the input audio signal via a respective microphone, and transmit the received input audio signal to the data processing system. A natural language processor (NLP) component of the data processing system can process the input audio request to identify the user request. The NLP component can machine-translate the audio query to a corresponding text and parse the generated text to identify one or more keywords. For example, the audio signal detected by the client device can include "Okay device, I need a ride from Taxi Service Company A to go to 1234 Main Street." The data processing system can identify the trigger keywords "I need," "to go to" and/or "ride," and determine a user request for a cab ride.

In some cases, the data processing system may generate an action data structure. The data processing system can generate, responsive to the user request, the action data structure based on the identified keywords, the determined user request, and/or other information. For example, if the user requests a taxi from Taxi Service Company A, the action data structure can include information to request a taxi service from Taxi Service Company A. The data processing system may select a template for Taxi Service Company A, and populate fields (e.g., user' identity information, pick-up location, pick-up time, destination, etc.) in the template with values to allow the Taxi Service Company A to send a taxi to the user to pick up the user and transport the user to the requested destination.

At step 110, the data processing system can provide a response component to the user request for presenting to the user responsive to the received input audio signal. The data processing system can generate the response component based on data received from a service provider responsive to the action data structure. For example, in the case of a request for a cab ride, the data processing can generate and send a ride request template to a computing device of

the cab service provider, and receive a response to the ride request template. The response can include information (e.g., taxi ID, driver's information, expected time of arrival, estimated cost, etc.) related to one or more potential rides. The data processing system can use the received information to generate the response component for sending to the user's client device. The data processing system can generate the response component as audio content, textual content, image content, video content, or a combination thereof. The data processing system may further select an ad (e.g., an ad related to another cab service provider) for presenting to the user. The data processing system may insert the selected ad within the response component or send it separately for presentation on the client device. For example, the selected ad may an audio ad that is inserted within an audio response component.

The data processing system can transmit the response component and/or the selected ad to the client device. The response component may include various elements with different formats. For example, the response component or the selected ad can include an audio signal (e.g., including information related to the ride(s) or an alternative service provider) and/or one or more visual elements (e.g., a text message or an image including a link of a webpage of a cab service provider). As such, different elements can be presented to the user via different interfaces or different client devices associated with the user. The user can interact with the response component and/or the selected ad received from the data processing system.

At step 115, the method 100 can include the data processing system (or a respective feedback monitoring component) receiving user interaction audio data responsive to presentation of the response component to the user. For example, the user can respond with a second input audio signal to confirm or select an offered ride or to ask further information. The user may also interact with the selected ad, for example, to request (e.g., via audio signal) more information

about an alternative service provider in the ad. The feedback monitoring component may also receive non-audio user interaction data (e.g., indication of user interacting with a visual content item, clicking a link therein, or executing an online transaction).

At step 120, the feedback monitoring component can measure (or determine) a quality of the conversation with the user based on received user interaction data (or user online activity data). Determining a quality of the conversation can include determining a quality level, quality metric, or quality score. For instance, the feedback monitoring component can determine a numeric quality score (e.g., 0 to 10 with 0 being lowest quality and 10 being highest quality, or vice versa), a letter grade (e.g., A to F with A being the best quality), a binary value (e.g., Yes/No; Good/Bad; 1/0; high/low), a rank, or a percentile value. The feedback monitoring component can assess the quality of the conversation based on communications (or user interaction data associated) with a plurality of client devices associated with the user.

The feedback monitoring component can assess the quality of the conversation with the user using various measuring techniques, heuristic techniques, policies, conditions, or tests. The quality of the conversation can include a quality of the communication channel used to transmit between the client device(s) of the user and the data processing system. The quality of the communication channel can include one or more indications of a signal-to-noise ratio (SNR), network signal level at a mobile client device, ambient noise level, latency, sound choppiness, echo, or dropped calls. The feedback monitoring model may estimate the noise in audio data received from the client device(s) and compute a SNR based on the estimated noise. The feedback monitoring component may receive indications of network signal level at mobile client device(s). The feedback monitoring component may track and/or record any call dropping events and/or latency (or travel time) for data exchanged between the client device(s) and the

data processing system. The feedback monitoring component may process audio signals received from the user client device(s) to detect any echo effects or choppiness effects. The feedback monitoring component may compare measured or detected values of the channel parameters (or characteristics) with respective predefined thresholds to assess the quality of the communication channel.

The feedback monitoring component can assess the quality of the conversation with the user based on performance parameters of the NLP component. For example the feedback monitoring component may monitor a confidence level of the speech recognition process applied to the input audio signals received from the client device(s). The feedback monitoring component may monitor time consumed by the data processing system to generate the response component starting from the time the input data signal is received at the data processing system. The feedback monitoring component may keep track of (or record) the number of times the user repeats an audio query. The feedback monitoring component may detect and keep track of any words, expressions, or tones (e.g., swears, obscenity, yelling, compliments, etc.) associated with audio signals received from the user that indicate the user's satisfaction level with the response component. The feedback monitoring component may keep track of (or record) events indicative of the user quitting the conversation with the virtual personal assistant, or the duration of the conversation. The feedback monitoring component may maintain states of the conversation and evaluate the time duration or the number response components generated (or number of input audio signals received) within each of the states. The feedback monitoring component may request explicit feedback (e.g., via an audio request and/or a request to fill a survey) from the user at the end of the conversation.

The feedback monitoring component can receive non-audio data indicative of user online

activity responsive to the response component provided by the data processing system. For example, if the user actuates a link provided by the data processing system, interacts with an ad or other content item, executes an online transaction, or creates an online account for a service on a given client device, the client device (or a computing device associated with a service provider) can report indications of such interactions to the feedback monitoring component. The feedback monitoring component may store the received indications in a data structure or database. The non-audio user interaction data may be received from a client device different than the client device that initiated the audio conversation or transmitted the input audio signals. For example, the input audio signals may be transmitted by a Google Home device whereas user online activity may occur on a smart phone, tablet, laptop, or desktop of the user.

The feedback monitoring component can generate one or more quality signals (or quality metrics) based on the collected quality data. For example, the feedback monitoring component may generate separate quality signals indicative of the qualities for the communication channel, the NLP component, and/or the content provided by the data processing system, respectively. The feedback monitoring component may generate each quality signal using a weighted sum of corresponding quality parameters. For example, the feedback monitoring component can determine the communication channel quality signal based on communication channel parameters or characteristics, such as SNR, latency, network signal level, detected echo effects, detected choppiness effects, call drops, or a combination thereof. The feedback monitoring component may generate a single quality signal that incorporates the qualities of the communication channel, the NLP component, and/or the content provided by the data processing system.

At step 125 the feedback monitoring component (or the data processing system) can

adjust a real-time content selection process based on the determined quality of the conversation. Adjusting the real-time content selection process can refer to adjusting a weight used to select content for presenting to the user, adding, removing or modifying a step in the content selection process, changing a format of the content provided for presentation to the user. For example, if content (e.g., a response component or an ad) presented to the user item led to a low content quality signal, the data processing system can adjust an attribute or parameter of the content data or a parameter of user preferences and/or user profile. Also, a low NLP component quality signal can lead the data processing system to add an extra step to ask the user to confirm the determined user intent, and/or adjust one or parameters of a speech recognition module of the data processing system. Upon determining a low communication channel quality signal, the data processing may decide to transmit visual response components (e.g., text messages) instead of audio response components to the user.

The process of monitoring conversation quality and adjusting selection process parameter(s) can be performed continually. The data processing system can be configured to take into account feedback data (or quality signals) associated with the most recent N conversations with the user, where N is an integer number, when adjusting the selection process parameter(s).

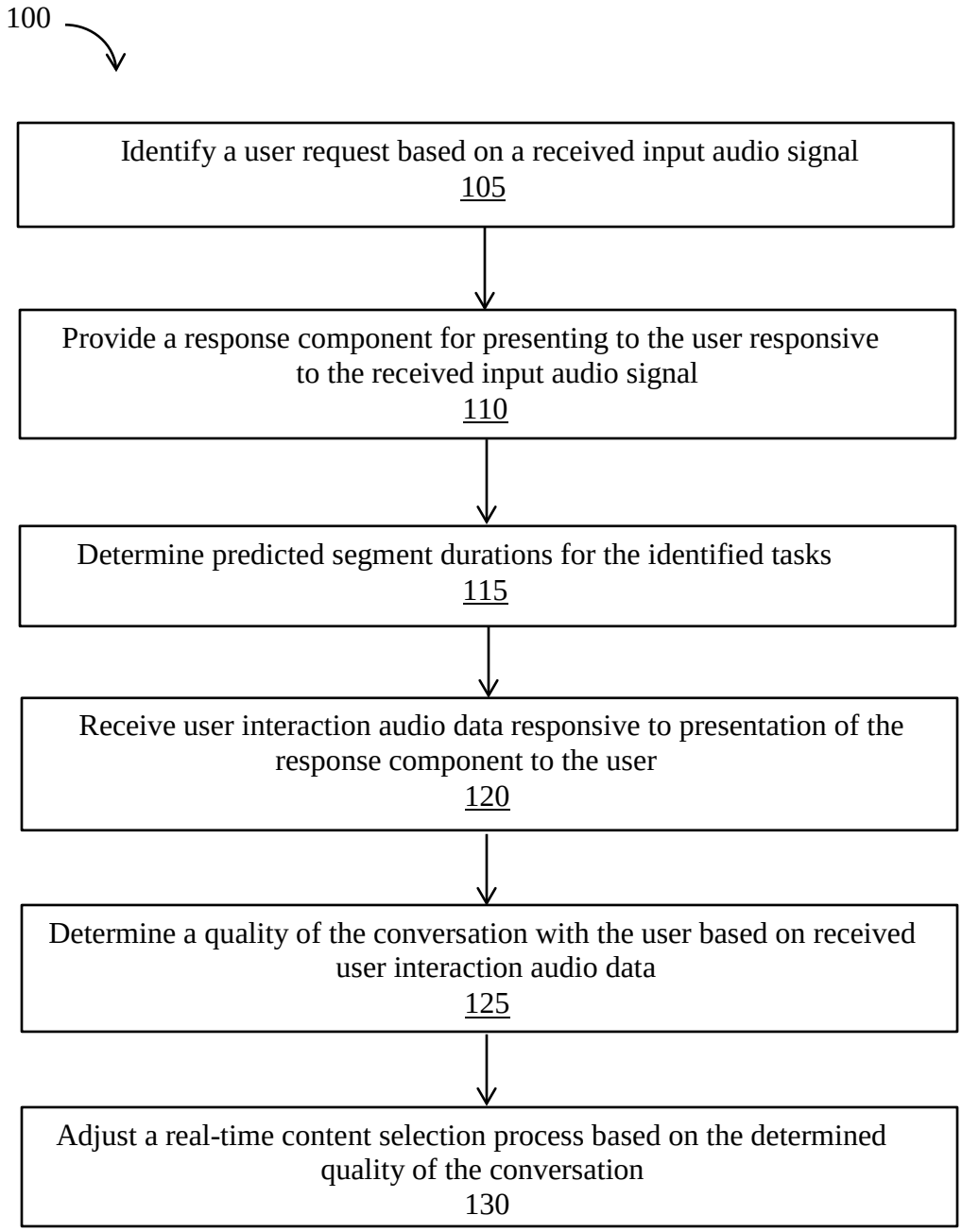
100
Identify a user request based on a received input audio signal
105
Provide a response component for presenting to the user responsive to the received input audio signal
110
Determine predicted segment durations for the identified tasks
115
Receive user interaction audio data responsive to presentation of the response component to the user
120
Determine a quality of the conversation with the user based on received user interaction audio data
125
Adjust a real-time content selection process based on the determined quality of the conversation
130

FIG. 1